The SCATE Prototype: A Smart Computer-Aided Translation Environment

Vincent Vandeghinste
KU Leuven
vincent@ccl.kuleuven.be

Sven Coppers
Hasselt University - tUL - imec
sven.coppers@uhasselt.be

Jan Van den Bergh
Hasselt University - tUL - imec
jan.vandenbergh@uhasselt.be

Tom Vanallemeersch
KU Leuven
tom@ccl.kuleuven.be

Bram Bulté
KU Leuven
bram.bulte@ccl.kuleuven.be

Ayla Rigouts Terryn
Ghent University
ayla.rigoutsterryn@ugent.be

Els Lefever
Ghent University
els.lefever@ugent.be

Iulianna van der Lek-Ciudin
KU Leuven
iulianna.vanderlekciudin@kuleuven.be

Karin Coninx
Hasselt University - tUL - imec
karin.coninx@uhasselt.be

Frieda Steurs
KU Leuven
frieda.steurs@.kuleuven.be

Abstract

We present the SCATE prototype: A Smart Computer-Aided Translation Environment, developed in the SCATE research project. Its user interface displays translation suggestions coming from different resources, in an intelligible and interactive way. It contains carefully designed representations that show relevant context to clarify why certain suggestions are given. In addition, several relationships between the source and the suggestions are made explicit so the user understands how a suggestion can be used in order to select the most appropriate one. Well-designed interaction techniques are included that improve the efficiency of the user interface. The suggestions are generated through different web services, such as fuzzy matching based on a translation memory (TM), machine translation (MT) and terminology extraction. MT and TM are combined using a pre-translation mechanism. A lookup mechanism highlights terms in the source segment that are available with their translation equivalents in the bilingual glossary.

This paper presents the interface and the underlying web services, and discusses preliminary evaluations of the interface and the pre-translation mechanism.

1 Introduction

We present a demonstration prototype of a computer-aided translation system that was built in the SCATE project (Smart Computer-Aided Translation Environment) (Vandeghinste et al. 2014). This project, which is currently in its final phase, investigates several aspects related to translation technology, such as the design of translators' user interfaces, the combination of machine translation (MT) and translation memory (TM), syntactic fuzzy matching, bilingual term extraction using parallel and comparable corpora, and confidence estimation of MT. The project is motivated by the fact that translators tend to have a limited trust in MT output, and translation environments provide a limited integration of resources.

The SCATE prototype consists of a carefully designed user interface that displays translation suggestions and terminology in an intelligible and interactive way. Translation suggestions are generated through a web service which integrates a TM system's fuzzy matching with MT. Terminology support is provided and terminology is automatically extracted from parallel corpora. Advanced autocompletion functionality allows to efficiently use the translation suggestions. While the SCATE prototype demo uses a medical corpus in a specific language pair (English-Dutch), the SCATE technology is sufficiently generic to be applicable to other domains and language pairs.

This paper is structured as follows. Section 2 describes the state-of-the-art in computer-aided translation environments. Section 3 describes the SCATE prototype. Section 4 provides details on a preliminary evaluation of the interface and of the combination of TM and MT. Section 5 discusses conclusions and future work.

2 State-of-the-art in computer-assisted translation

Computer-assisted translation (CAT) tools have been commercialised since the late 1990s, triggering new business models and greatly influencing the translation and localisation processes, and the way translators work. Users can perform basic project management tasks, create and maintain TM and terminology databases, query MT engines and online databases directly from the translation editor, automatically extract terms from reference materials, align parallel corpora, and use automatic quality control checks on the target document to detect various types of errors. Moreover, cloud-based systems have made collaboration much easier as an entire team can work on the same text simultaneously in real time, leaving comments, sharing and updating resources instantly.

Despite the wide range of functionalities and possibilities, CAT tools are not used to their true potential either because of usability issues or because the integration of various technologies (TM, MT, term bases) is not yet optimal (Ehrensberger-Dow and O'Brien, 2015; Zaretskaya, 2015; Krüger, 2016; Moorkens and O'Brien, 2016). Moreover, translators have not fully adopted MT as an aid because they do not trust the quality of the commercial MT engines (Van den Bergh et al., 2015; Cadwell et al., 2017). We briefly review current commercial translation environments according to two criteria: usability and extent of integration of different resources (terminology, TM and MT).

2.1 Usability

The user interface of CAT tools typically provides access to resources such as translation memories (TM), machine translation (MT) and terminology databases (TB). Tools differ in the way resources are made available, more specifically in terms of the visual proximity of suggestions, information provided on the origin of (parts of) a translation suggestion, and options to facilitate the reuse of sub-segments from TM or MT.

With regard to the visual ***proximity of suggestions***, these are ideally displayed on a single screen, together with the surrounding context of the segment being translated, as translators like to have all the information at their fingertips (Lagoudaki, 2009). Different approaches have been adopted: some tools offer a limited amount of suggestions close to the active working area, while others offer each of these resources in dedicated subwindows. For instance, MateCat[1] shows relevant resources in a tabbed interface immediately below the active working area, while Lilt[2] shows one suggestion from either TM or MT in the same

[1] https://www.matecat.com/
[2] https://lilt.com/

field. The second approach, as exemplified by SDL Trados Studio[3] and WordFast,[4] allows to select the type of resource to be displayed, thus limiting the variety of information that is close to the active working area. Moorkens and O'Brien (2016) confirms that there are proponents for both approaches. In SCATE, we follow the first approach as it eliminates visual focus shifts (see Section 3).

Most CAT tools offer limited information on the ***origin of (parts of) a translation suggestion***. The focus is mainly on highlighting the differences between the text to translate and matches from the TM (including match percentages). For MT, most of the time no justification is provided; typically, MT is used as a black box. Teixeira (2014:171) shows that metadata can help translators make well-informed decisions. He concludes that metadata "helps translators adapt their translation strategies more easily according to the suggestion type". Moorkens and O'Brien (2016) indicates that translators like information about the provenance of the MT suggestions and estimation of their quality. In the context of post-editing, Viera and Specia (2011) argues that translators value on-the-fly highlighting of word alignment in order to keep the connection between source and target text. In other words, it appears useful to explicitly link parts of a source sentence with parts of the translation suggestion. As discussed in Section 3, the SCATE prototype is strongly focused on providing visual aids that explain the origin of translation suggestions and their link with the source text.

As for ***reuse of sub-segments from TM or MT***, we point to recent user research, including surveys and field studies (Van den Bergh et al., 2015; Moorkens and O'Brien, 2016), that investigates the interaction between machines and translators. One conclusion is that translators value improved TM-MT integration methods (e.g. copy/paste, drag-and-drop within editor). Reuse of sub-segments is also possible through ***interactive translation prediction*** (ITP) (Koehn and Haddow, 2009; Sanchis-Trilles et al., 2014; Torregrosa et al., 2017). This is a form of human-computer interaction in which users are presented, as they type, with translation suggestions from all available resources. Suggestions are displayed either in a drop-down list or directly under the target segment. Commercial translation software developers have implemented this technology in different ways and use different terms to refer to it: *predictive typing*, *AutoSuggest*, *Autocomplete*, *Autowrite*. Research has shown that translators prefer ITP to classical post-editing because it minimizes the number of keystrokes and thus increases productivity (Koehn and Haddow, 2009; Sanchis-Trilles et al., 2014; Zaretskaya, 2015).

2.2 Integration

Translation environments typically include functionalities for terminology management and support for MT. Terminology management mostly consists of basic features to retrieve, save, search, import/export, and maintain terms and term bases. MT integration takes place either via plugins or by combining MT with various other linguistic resources (TM, TB).

While some translation environments include a tool that can be used to ***extract potential terms*** from TM (automatic term extractor), research conducted within SCATE shows that term extraction has not yet become a standard practice in the translator's workflow (Steurs et al., 2016; Van den Bergh et al., 2015), leading translators to mainly rely on their TMs and concordance features. Whereas a hybrid approach (combining both linguistic and statistical

[3] http://www.sdl.com/software-and-services/translation-software/sdl-trados-studio/
[4] http://www.wordfast.com/

methods) may be the best method for preparing terminology collections in commercial environments (Warburton, 2015), most translation environments are still limited to monolingual statistical term extraction that often produces either too much “noise” (too many general lexicon words) or “silence” (real terms that are ignored). Moreover, there is still a lack of ***integration of terminology in translation editors***. In order to tackle the problem of integration, the Lilt tool,[5] to give an example, combines the glossary with a concordance feature and updates both resources while the translator works. In the SCATE prototype, we approach both of the above issues by incorporating a bilingual term extractor and by smoothly integrating bilingual terminology in the translation editor.

Translation environments usually ***integrate MT and TM*** in a rather trivial way. They either offer the translation of a fuzzy match (given some threshold) or an MT suggestion. A growing body of research has explored different ways of combining information coming from TMs and MT. An MT system can be constrained to the use of relevant parts of a fuzzy match (Zhechev and van Genabith, 2010), for example by adding XML markup to MT input (Koehn and Senellart, 2010). Other methods have focused on augmenting the translation table of a phrase-based MT system with aligned sub-segments from a retrieved TM match (Biçici and Dymetman, 2008). Alternatively, information from the fuzzy matches can also be integrated in the MT system itself (Li et al., 2017). In the SCATE prototype, we opt for an approach which is based on XML markup and relatively straightforward to implement (see Section 3).

The next section describes the SCATE prototype in more detail, focusing on the main user interface and the integration of the different technologies.

3 The SCATE prototype

The SCATE prototype is a web-based translation environment, built through a user-centered development approach. Figure 1 presents an overview of the user interface. We focus on usability and on interaction techniques to integrate various translation technologies. The contributions within the SCATE user interface explain *how* different translation suggestions are generated, *why* they might be useful to the translator and *which* relationships exist between them. The aim is to support the translator’s decision-making process during the selection of a translation suggestion. For the demo we use the English-Dutch part of the medical corpus EMEA (Tiedemann, 2007), which contains about 300,000 sentence pairs. This resource was used to train the MT, as a TM, and as resource for bilingual term extraction.

3.1 Usability

The SCATE prototype includes four different translation aids: (1) matches from the TM, (2) hybrid MT, (3) alternatives for the selected term and (4) an autocomplete feature to predict the rest of a word or word group. We developed a web service that accepts the sentence to translate and provides fuzzy matches (Figure 1.E) and MT output (Figure 1.C).

[5] https://lilt.com/kb/translators/lexicon

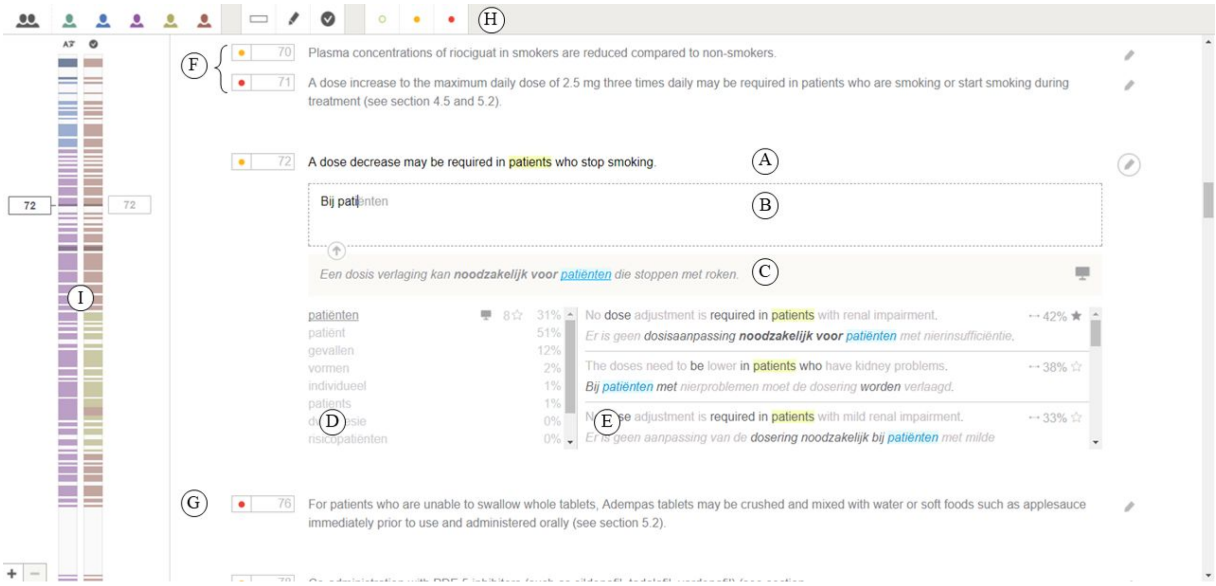

Figure 1. Overview of the SCATE user interface. (A) Sentence to translate, (B) edit field, (C) hybrid MT, (D) translation alternatives and (E) TM matches. At all times, the preceding (F) and subsequent sentences (G) remain visible. Vertical bars (I) can visualise active filters (H) such as difficulty, responsibilities and progress.

To support better decision-making about the use of translation suggestions, the user interface focuses on ***intelligibility***. This focus is key to clarify the behaviour of the complex algorithms behind the translation suggestions. The algorithms that find matches (Figure 1.E) in the TM are made intelligible to the translator with icons that depict the used matching metric, with scores representing the level of similarity, and by highlighting parts in the matches that are potentially useful. Existing CAT tools, such as MateCat, at most highlight differences instead of similarities. In the SCATE prototype, partial matches that are often translated by the same group of words are used as pre-translations by the MT engine (Figure 1.C). To make this clear to translators, pre-translations are shown in bold in both the matches and the MT suggestion. On the left side of the matches, potential term translation options, aggregated from TB, TM and MT are listed (Figure 1.D), each with a metric informing the translator about its usefulness. For MT and TM, an absolute value is shown to represent how often an option occurs in the TM matches. For the TB, we show relative frequency. For ITP, all options are considered and can be manually added to the translation. To further enhance intelligibility, occurrences of these options in the matches are automatically highlighted.

In addition to multiple preceding (Figure 1.F) and subsequent sentences (Figure 1.G), all translation aids remain visible at all times, eliminating the visual focus shifts typically required in other CAT tools (Ehrensberger-Dow et al., 2014; Lagoudaki, 2009). Furthermore, a simultaneous exploration of up to four different kinds of relationships between various sorts of suggestions is supported when typing in the editing field (Figure 1.B) or when hovering the mouse cursor over a word in any sentence. (1) Words in the sentence that belong to the same word group are highlighted in the same colour. (2) Translations of the hovered word are highlighted within the TM and MT. Words in the source language appear in yellow, whereas words in the target language appear in blue. (3) Synonyms and alternative translations of the word appear within the matches and MT suggestion in the same colour as the word itself. An overview of all synonyms is shown in the alternatives list (Figure 1.D). This overview works in the inverse direction as well: by hovering over an alternative in the list, (4) occurrences of the alternative are highlighted in the matches from the TM. When the first occurrence of the alternative is not within the viewport (the part of a scrollable window currently visible), the panel with matches will automatically scroll.

3.2 Integration

The SCATE prototype combines ***web services connecting to a TM and a phrase-based statistical MT system***, Moses (Koehn et al., 2007). TM matches are retrieved using three metrics: Levenshtein distance (Levenshtein, 1966), METEOR (Lavie and Agarwal, 2007) and *shared partial subtree matching*, a measure based on syntactic similarity (Vanallemeersch and Vandeghinste, 2015). For each sentence, the N best matches from the TM (according to the fuzzy match score) are stored in a reduced TM subset, together with information on the match score, rank, used fuzzy metric, and part-of-speech (POS) sequence of both the source and target sentence. Two 'sliders' can be set by the user: the TM slider and the MT slider. Matches with a score higher than the TM slider are directly used as final translation, and sentences which have no fuzzy match at all or no fuzzy match that scores higher than the MT-slider are sent straight to the Moses SMT system (as illustrated in Figure 2).

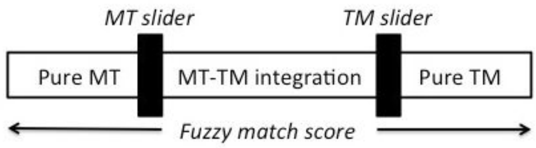

Figure 2. System of 'sliders' for the selection of MT, TM-MT integration or TM.

To produce hybrid TM-MT suggestions, the MT system is constrained to use certain word sequences (or pre-translations) extracted from the TM matches. Initially, a four-stage alignment procedure is followed for each triplet of input sentence, TM source sentence and TM target sentence (see Figure 3). Step 1 identifies the overlapping spans between the input sentence and TM source sentence. Step 2 aligns the TM source sentence with the TM target sentence at the word level using the automatic word alignment and lexical probabilities derived by GIZA++ (Och and Ney, 2003) and Moses. Step 3 identifies consistently aligned spans[6] for the TM source and target sentence, using the grow-diag-final heuristic (Koehn, 2009), and consistently aligned sub-spans of these spans are identified based on the same criteria. Finally, step 4 couples the consistently aligned spans between TM source and target to overlapping spans in the input sentence.

The extracted spans are subsequently filtered (based on criteria such as minimum span length, occurrence of at least one content word, and percentage of aligned words), weighted (taking into account span length, span frequency across TM matches, and fuzzy match score of the strongest match in which the aligned span occurs) and ranked. The best ranked non-overlapping spans are added to the input sentence using XML markup, and these augmented input sentences are sent to Moses.

[6] Pairs of spans in the source and target language in which words are not aligned with words outside the spans.

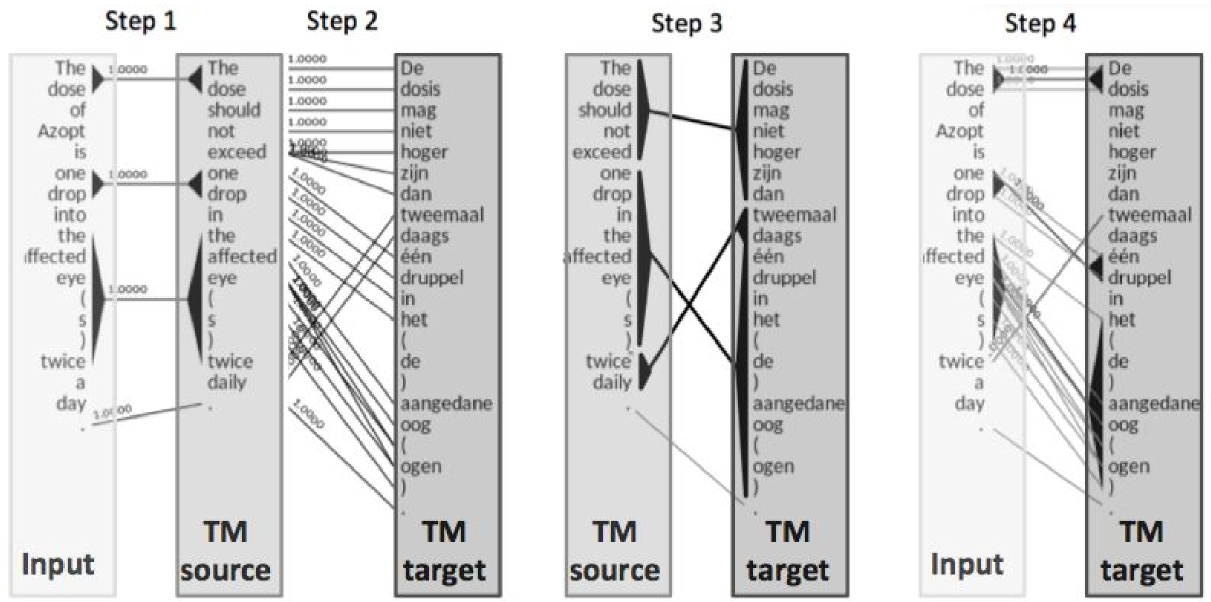

Figure 3. Illustration of alignment procedure. Step 1: identification of overlapping spans in input and TM source. Step 2: word-level alignment between TM source and target. Step 3: finding consistently aligned spans. Step 4: coupling aligned spans in TM target to input.

As for the integration of terminology, we generated a ***bilingual TB*** offline using *TExSIS* (Macken et al., 2013), a ***hybrid terminology extraction tool*** that uses POS patterns to obtain a preliminary list of candidate terms, which is subsequently filtered statistically. The list of alternatives for the selected term (Figure 2.D) aggregates suggestions from MT and TM on the basis of word alignment, and from the bilingual TB.

When generating the bilingual TB, we restrict our search to nouns, noun phrases and adjectives, and ensure long multiword terms (MWTs), such as *cholangiocarcinoma of the extrahepatic bile ducts,* are not omitted. As this linguistic strategy overgenerates since it extracts every occurrence of all valid linguistic patterns (i.e. all nouns, all adjectives, all noun+noun etc.), we apply statistical detection of terms, using the two main principles of *termhood* and *unithood*. Termhood indicates the specificity of lexical entries to a certain domain, and is calculated by comparing the relative frequencies of the candidate term in the domain-specific corpus with a general language corpus. Unithood only applies to MWTs and is based on the idea that MWTs are more than the sum of the meaning of the different parts and that the different parts are strongly connected. Therefore, unithood is computed by comparing the frequencies of cooccurrence of the parts with the expected cooccurrence based on the relative frequencies of each of the parts. While both principles are effective, some terms may go undetected, for instance when they are too rare or new, meaning that the term frequency is too low to get significant results for any statistical measure. Due to the Zipfian distribution of language (the long tail distribution of rare words), non-terms may be extracted as well, such as idiomatic phrases (e.g. *significant part*) or (multi)words that are not domain-specific (e.g. *guitar players* was extracted from the medical corpus).

Based on the monolingual lists of term candidates and the sentence-aligned input corpus, source and target term candidates are linked to each other, in order to generate the bilingual TB. To this end, word alignment is performed on the corpus using the GIZA++ word alignment toolkit. Based on these word alignments, each candidate term in the source language is linked to a candidate term in the target language. These results are filtered by comparing the frequency of the source language candidate term with the number of times it is linked to the target language term according to the word alignment. If this results in a value of less than 20%, the target language term is discarded as a translation suggestion. Discarded suggestions include partial translations (e.g. *medication* - *diureticummedicatie*) and wrong spellings. Correct suggestions may be discarded because the source and target term have a different POS tag (e.g. *x-ray* - *radiologie, 'radiology'*). On the other hand, incorrect suggestions may be retained: for instance, one term may be the hyponym or hypernym of the other (e.g. *patient* - *kind, 'child'*), or terms may be only loosely related (e.g. *treatment* - *medicatie, 'medication'*).

4 Evaluation

At various stages of the SCATE project, we involved professional translators and translation experts in the design of techniques and interfaces, and in preliminary evaluations. We carried out a formative study in which 8 participants used two versions of the SCATE prototype to translate a text. Both versions provided the same translation suggestions, but in the first version, these suggestions were presented without the intelligibility features described in Section 3, to measure their impact on the user experience. The results show that making more contextual information available has a positive impact. The general usability increased slightly from 71.6 to 76.6 (above average), as measured by the SUS scoring method (Brooke, 1996). Judging by the overall comments, participants highly appreciated the intelligible version. More specifically, professional translators value the fact that match scores are indicated, that words in the TM which match the sentence to translate are highlighted, and that relationships between suggestions are made explicit through visual marks. These features help them to better understand why a translation suggestion might be useful or not, while not being perceived as distracting. Contrary to our expectations, displaying more meta-information is not always desired by our participants. We point out that the quality of the suggestions is always more important than making them more understandable.

As for in vitro testing of the integration of TM and MT, we carried out preliminary tests on three TMs (EMEA, DGT[7] and a TM provided to us by a software development company[8]) to evaluate the quality of the hybrid TM-MT suggestions. For each of the datasets, three automated evaluation metrics (BLEU, METEOR and TER) indicated a significant increase in translation quality compared to 'pure' MT output. Additionally, qualitative spot checks by translators revealed that in a majority of cases the hybrid suggestions proved to be better than the pure MT output in terms of grammaticality and/or fluency, or provided interesting translation alternatives.

5 Conclusions and future work

We presented an innovative prototype CAT system that was built in the SCATE project. The prototype combines different types of translation suggestions into a carefully designed user interface and makes the suggestions available through ITP. The visualisations remain compact and are presented close to the sentence to be translated. We apply bilingual (instead of monolingual) term extraction, combine statistics with linguistic patterns during extraction, and access MT as a glass box: internal information from the phrase-based MT system is used

[7] Subset of 1.7 million sentences (Steinberger et al., 2013).
[8] 150 000 sentences. Provided through a confidentiality agreement.

to produce hybrid MT output and to visualise links between the sentence to translate, the MT output and fuzzy matches.

Preliminary evaluation of the prototype shows that providing more metadata in an intelligible and interactive manner is not perceived as distracting and helps translators to decide on the best translation suggestions. *In vitro* evaluation of the hybrid MT output has shown that it produces more useful translation suggestions than pure MT. In addition to the increased quality of the MT output, the highlighting of pre-translations taken directly from the TM has the potential of increasing translators' confidence in MT output. This, however, needs to be further studied.

Current and future work includes the integration of a quality estimation metric for MT (Tezcan et al., 2017), options to configure the translation workflow, as well as support for terminology extraction from comparable corpora (Bowker, 2003; Delpech, 2014). With regard to TM-MT integration, we intend to include functionality for automated fuzzy match repair (Ortega et al., 2016) and perform in-depth tests of syntactic fuzzy matching. More specific evaluations which focus on the impact of intelligibility on the user experience and performance are ongoing. Since the techniques developed in SCATE are generic, we plan to perform tests with other language pairs and domains. Finally, we intend to perform a comparative evaluation of the SCATE prototype with another state-of-the-art tool.

Acknowledgements

The SCATE project is funded by the Flemish agency for Innovation and Technology (IWT), under project number 130041.

References

Biçici, Ergun and Marc Dymetman. 2008. Dynamic translation memory: using statistical machine translation to improve translation memory fuzzy matches. *CICLing*, Haifa, Israel, pages 454–465.

Bowker, Lynne. 2002. Working Together: A Collaborative Approach to DIY Corpora. In *The First International Workshop on Language Resources for Translation Work and Research*, pages 29–36.

Brooke, John. 1996. SUS: A quick and dirty usability scale. In P. W. Jordan et al. (Eds.), *Usability Evaluation in Industry*, 189 (194). Taylor and Francis, London, pages 4–7.

Cadwell, Patrick, Sharon O'Brien, and Carlos S. C. Teixeira. 2017. Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators. In *Perspectives Studies in Translatology*, pages 1–21.

Delpech, Estelle. 2014. Leveraging Comparable Corpora for Computer-assisted Translation. In *Comparable Corpora and Computer-Assisted Translation*. John Wiley & Sons, Inc., Hoboken, NJ, USA, pages 1–39.

Ehrensberger-Dow, Maureen and Gary Massey. 2014. Cognitive ergonomic issues in professional translation. In: Schwieter, John W. & Aline Ferreira (eds), *The Development of Translation Competence: Theories and Methodologies from Psycholinguistics and Cognitive Science*. Newcastle upon Tyne: Cambridge Scholars Publishing, pages 58–86.

Ehrensberger-Dow, Maureen and Sharon O'Brien. 2015. Ergonomics of the translation workplace: Potential for cognitive friction. In *Special Issue of Translation Spaces*, 4 (1), pages 98–118.

Koehn, Philipp. 2009. *Statistical machine translation.* Cambridge: Cambridge University Press.

Koehn, Philipp and Barry Haddow. 2009. Interactive Assistance to Human Translators Using Statistical Machine Translation Methods. In *Machine Translation Summit XII* , pages 73–80.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th ACL*, demonstration session, pages 177–180.

Koehn, Philipp and Jean Senellart. 2010. Convergence of translation memory and statistical machine translation. In *Proceedings of the 2nd Joint EM+/CNGL Workshop* “Bringing MT to the User: Research on Integrating MT in the Translation Industry”(JEC '10), pages 21–31.

Krüger, Ralph. 2016. Contextualising Computer-Assisted Translation Tools and Modelling Their Usability. In *Trans-Kom - Journal of Translation and Technical Communication Research*, 9(1), pages 114–148.

Lagoudaki, Elina 2009. Translation editing environments. In *MT Summit XII*: Workshop Beyond Translation Memories.

Lavie, Alon and Abhaya Agarwal. 2007. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of WMT*, pages 228–231.

Levenshtein, Vladimir. 1966. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. In *Soviet Physics-Doklady*, 10, pages 707–710.

Li, Liangyou, Carla P. Escartín, Andy Way, and Qun Liu. 2017. Combining translation memories and statistical machine translation using sparse features. In *Machine Translation*, 30(3), pages 183–202.

Macken, Lieve, Els Lefever, and Véronique Hoste. 2013. TExSIS: Bilingual Terminology Extraction from Parallel Corpora Using Chunk-based Alignment. In *Terminology*, 19 (1), John Benjamins Publishing Company, Amsterdam, Netherlands, pages 1–30.

Moorkens, Joss and Sharon O’Brien. 2016. Assessing user interface needs of post-editors of machine translation. In *Human Issues in Translation Technology: The IATIS Yearbook*, pages 109–130.

Och, Franz J. and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. In *Computational Linguistics*, 29(1), pages 19–51.

Ortega, John E., Felipe Sánchez-Martınez, and Mikel L. Forcada. 2016. Fuzzy-match repair using black-box machine translation systems: what can be expected? In *Proceedings of AMTA*, Vol. 1, pages 27–39.

Sanchis-Trilles, Germán, Vicent Alabau, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Ulrich Germann, Jesús González-Rubio, Robin L. Hill, Philipp Koehn, Luis A. Leiva, Bartolomé Mesa-Lao, Daniel Ortiz-Martínez, Herve Saint-Amand, Chara Tsoukala, Enrique Vidal. 2014. Interactive translation prediction versus conventional post-editing in practice: a study with the CasMaCat workbench. In *Machine Translation*, *28*(3-4), pages 217–235.

Steinberger, Ralf, Andreas Eisele, Szymon Klocek, Spyridon Pilos, and Patrick Schlüter. 2013. DGT-TM: A freely available translation memory in 22 languages. In *Proceedings of LREC*, pages 454–459.

Steurs, Frieda, Iulianna van der Lek-Ciudin, and Tom Vanallemeersch. 2016. How translators work in real-life: SCATE observations. In *Translating for Europe Forum*. Brussels, Belgium, 27-28 October 2016.

Teixeira, Carlos S. C. 2014. Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories. In *Third Workshop on Post-Editing Technology and Practice*, pages 45–59.

Tezcan, Arda, Véronique Hoste, Bart Desmet, and Lieve Macken. 2015. UGENT-LT3 SCATE System for Machine Translation Quality Estimation. In *Proceedings of WMT*, pages 353–360.

Tiedemann, Jörg. 2009. News from OPUS-A collection of multilingual parallel corpora with tools and interfaces. In *RANLP*, pages 237–248.

Torregrosa, Daniel, Juan Antonio Pérez-Ortiz, and Mikel L. Forcada. 2017. Comparative Human and Automatic Evaluation of Glass-Box and Black-Box Approaches to Interactive Translation Prediction. *The Prague Bulletin of Mathematical Linguistics*, 108(1), pages 97–108.

Vanallemeersch, Tom and Vincent Vandeghinste. 2015. Assessing linguistically aware fuzzy matching in translation memories. In *Proceedings of EAMT*, Antalya, Turkey, pages 153–160.

Vandeghinste, Vincent, Tom Vanallemeersch, Frank Van Eynde, Lieve Macken, Els Lefever, Véronique Hoste, Marie-Francine Moens, Joris Pelemans, Patrick Wambacq, Mieke Haesen, Karin Coninx, and Ken De Wachter. 2014. Smart Computer Aided Translation Environment. In *Proceedings of EAMT*, Dubrovnik, Croatia, page 135.

Van den Bergh, Jan, Eva Geurts, Donald Degraen, Mieke Haesen, Iulianna van der Lek-Ciudin, and Karin Coninx. 2015. Recommendations for Translation Environments to Improve Translators' workflows. In *Translating and the Computer 37*, Asling, London, pages 106–119.

Vieira, Lucas Nunes and Lucia Specia. 2011. A review of translation tools from a post-editing perspective. In *3rd joint EM+/CNGL Workshop bringing MT to the user: Research meets translators* (JEC), pages 33–42.

Warburton, Kara. 2015. Managing terminology in commercial environments. In *Handbook of Terminology*, Volume 1, John Benjamins Publishing Company, pages 359–391.

Zaretskaya, Anna. 2015. The use of machine translation among professional translators. In *Proceedings of the EXPERT Scientific and Technological Workshop*, pages 1–12..

Zhechev, Ventsislav and Josef van Genabith. 2010. Seeding statistical machine translation with translation memory output through tree-based structural alignment. In *Proceedings of SSST*, pages 43–51.